Project Report:

Binary Classificationon ‘Customer_Churn’ using Keras

Submitted by -Sekhar Chandra Padhi

1. Introduction

This project aims to analyze customer churn for the telecom company "Leo" and build predictive models to identify customers at risk of churning. Customer churn, or customer attrition, is the loss of clients or customers.

2. Data Source

The project uses the "Telco Customer Churn" dataset from Kaggle. The dataset contains information about customers, including demographics, services subscribed to, contract details, and churn status.

3. Data Manipulation

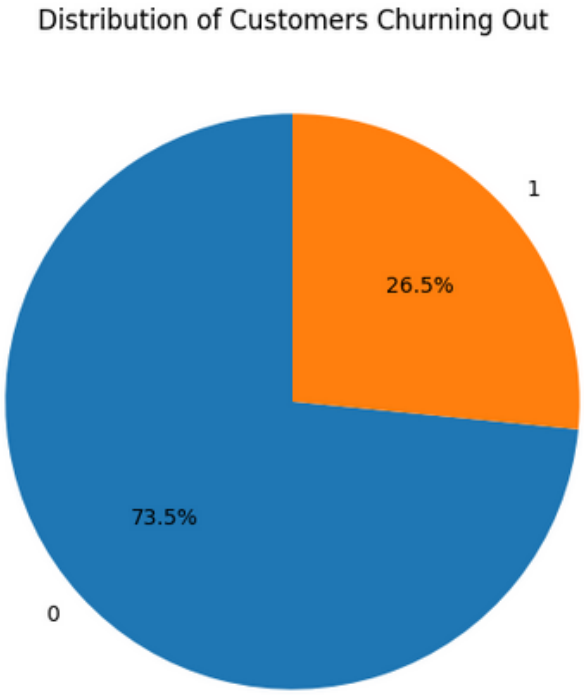
- The dataset was loaded and inspected using Pandas.
- Data cleaning and preprocessing steps were performed, such as converting data types and handling missing values.

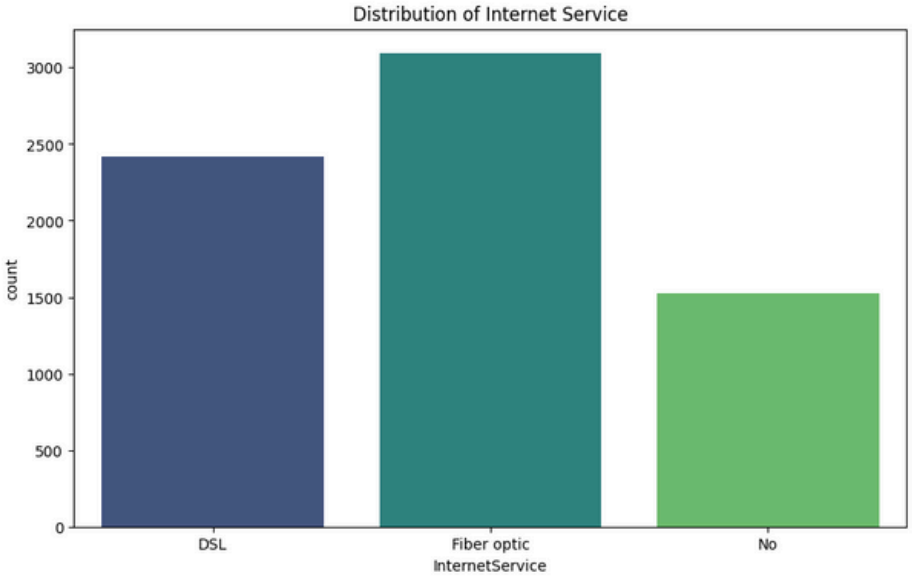
The following data manipulation tasks were performed:

1. Calculating the total number of male customers.
2. Finding the total number of customers with DSL internet service.
3. Extracting specific customer segments based on demographics and payment methods.
4. Filtering customers based on tenure and total charges.

4. Data Visualization

- A pie chart was created to visualize the distribution of customer churn.
- A bar plot was generated to show the distribution of internet service types among customers.



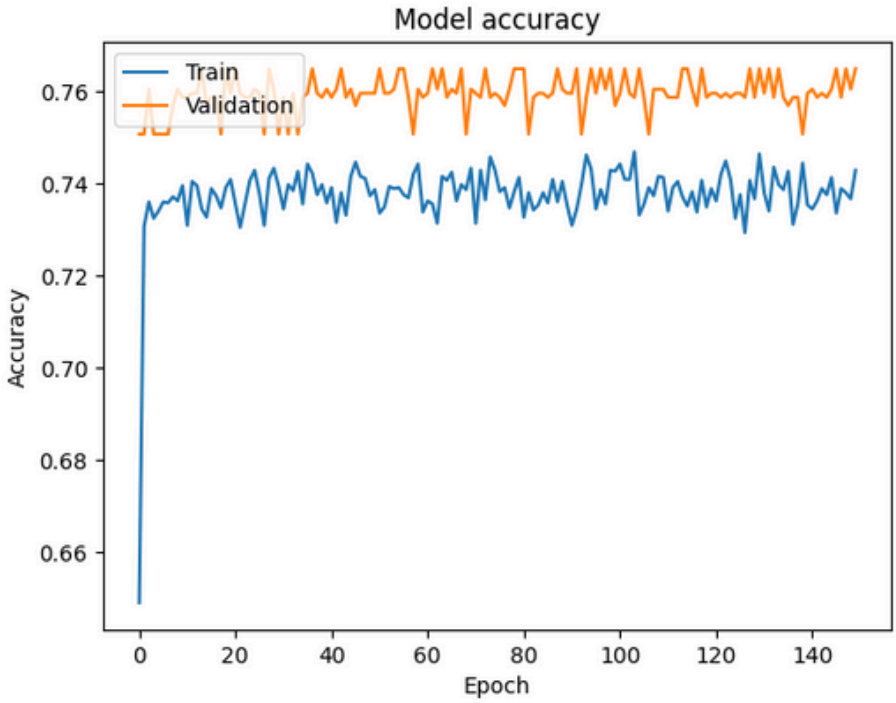


5. Model Building

Three sequential models were built using Keras to predict customer churn:

Model 1:

- Features: Tenure
- Target: Churn
- Architecture:
  Input layer: 12 nodes, ReLU activation
  Hidden layer: 8 nodes, ReLU activation
  Output layer: 1 node, Sigmoid activation
- Optimizer: Adam
- Epochs: 150

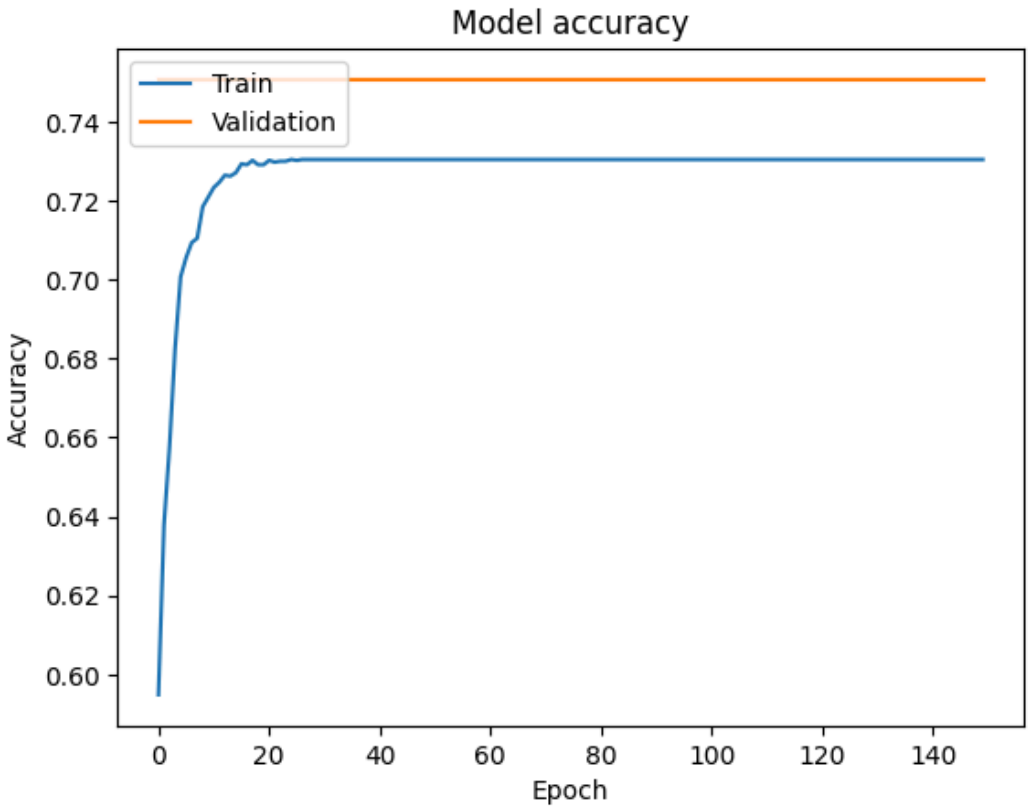


Accuracy: 0.75

Confusion Matrix:
[[950 86]
[257 116]]

Model 2:

- Similar to Model 1 but with dropout layers added after the input and hidden layers to prevent overfitting.
- Dropout values: 0.3 after the input layer and 0.2 after the hidden layer.



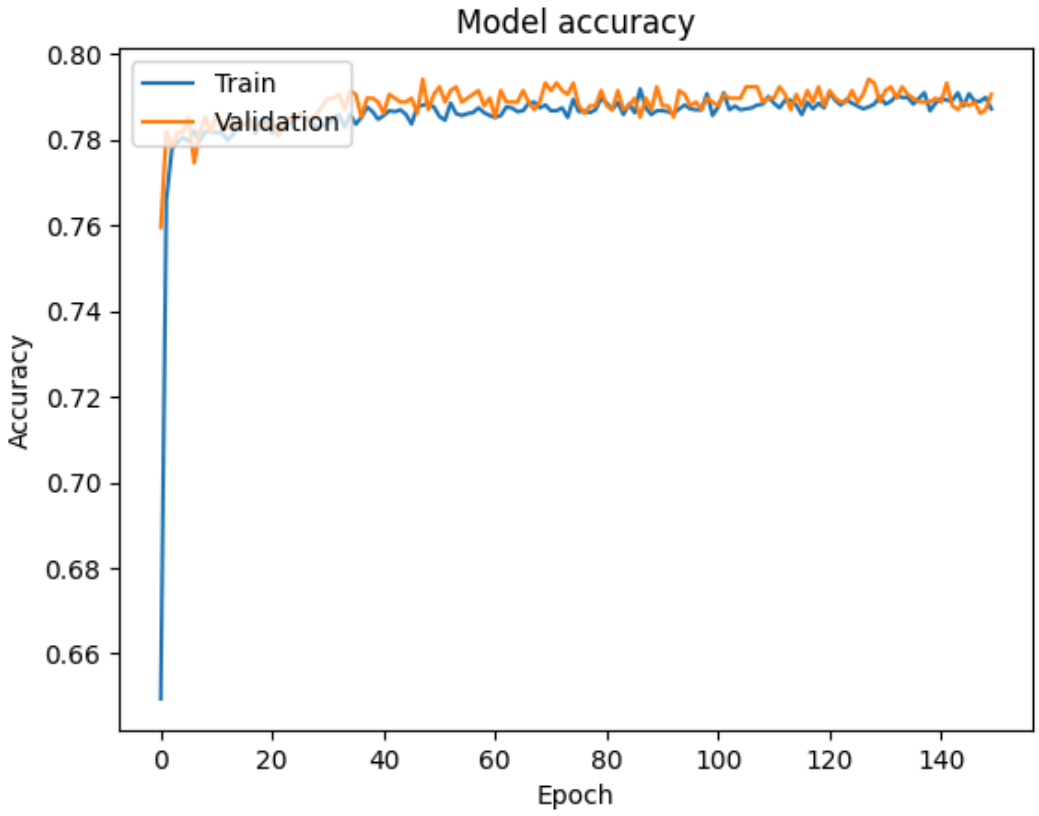
Accuracy: 0.73

Confusion Matrix:
[[1036 0]
[373 0]]

Model 3:

- Features: Tenure, Monthly Charges, Total Charges
- Target: Churn
- Architecture:

  Input layer: 12 nodes, ReLU activation
  Hidden layer: 8 nodes, ReLU activation
  Output layer: 1 node, Sigmoid activation
- Optimizer: Adam
- Epochs: 150



Accuracy: 0.79

Confusion Matrix:
[[955 81]
[203 170]]

7. Conclusion

The project successfully analyzed the Telco Customer Churn dataset and built predictive models using Keras. The models demonstrated varying levels of accuracy in predicting customer churn. Further improvements can be explored by incorporating more features, experimenting with different model architectures, and hyperparameter tuning.

8. Future Work

- Explore feature engineering to create new features that might improve model performance.
- Experiment with different deep learning architectures, such as recurrent neural networks (RNNs) for better prediction accuracy.
- Fine-tune the hyperparameters of the models to optimize their performance.
- Deploy the best-performing model for real-time churn prediction and implement strategies to retain customers at risk of churning.